



NSCF Georeferencing Protocol V2.0

Filename and Version:	NSCF Georeferencing Protocol V2.0
Release Status:	Version for publication in NSCF Manual
Version	2.0
Author:	Ian Engelbrecht
Author capacity	NSCF Data Coordinator
Contributors	Thus far: Albe Bosman, Hester Steyns, Burgert Muller, Christiaan Steenkamp, Adriaan Jordaan. Many others still be added but see manual acknowledgements
Date:	2021/05/29
Source:	NSCF Data Working Group
Reason for change/version update:	Minor changes and removed comments
Confidentiality Statement	<i>All documents and products related to the NSCF project and under development should be treated as confidential, and should only be shared with the project team and relevant and intended users who are part of the NSCF network. Final versions may be made publicly available if explicitly stated.</i>

Georeferencing tools:	2
Terrestrial and freshwater georeferencing:	2
Marine georeferencing:	3
What constitutes a georeference	3
Coordinate formats	4
Georeference uncertainty	4
Protocol	6
Locating features	6
The locality value	6
Districts, regions, cities, large mountain ranges, and uncertain localities	7
Suburbs	7
Towns	8
Named places other than towns	8
Named place with distance and direction from another named place	10
Distance and direction from a named place only	10
Farms and protected areas	13
Mountains, hillsides and mountain passes	14

Localities with coordinates	15
Quarter degree squares (QDSs)	16
Using habitat information and elevation	17
Ambiguous or contradictory localities	17
Georeference sources	18
Georeference remarks	18
Completing the remaining fields	18
Common abbreviations in location strings	18

1. Georeferencing tools:

A comprehensive suite of software tools is recommended for the georeferencing process. The tools described below can be downloaded from the NSCF website under Resources >> Standards and protocols. Training videos on the use of these tools and this protocol are available at

1.1. Terrestrial and freshwater georeferencing:

- QGIS software with the following layers and plugins
 - African countries
 - Administrative boundaries for southern African countries (provinces, districts, etc.)
 - For South Africa:
 - Magisterial Districts
 - Farms and farm centroids
 - Quarter degree squares
 - Protected Areas
 - 1:250 000 and 1:50 000 topocadastral maps
 - Rivers
 - Google Maps background layers (standard, terrain and satellite)
 - QGIS Plugins:
 - SANBI Gazetteer
 - Lat Lon Tools
- The following websites open for working with:
 - Nominatum (nominatim.openstreetmap.org)
 - Google Maps
 - Georeference Calculator
(<http://georeferencing.org/georefccalculator/gci3/source/gci3.html>)
 - Google Search (for web searches)
 - Geolocate (<https://www.geo-locate.org/web/WebGeoref.aspx>)
 - GeoNames (<https://www.geonames.org/>)

1.2. Marine georeferencing:

- QGIS software with the following layers and plugins:

For coordinates obtained from sources outside the QGIS project provided by the NSCF or Google Maps, make sure to copy and paste the coordinates into QGIS, check that they are in the right place (they very often need to be adjusted) and select the coordinates to use for the georeference again using Lat Long Tools in QGIS.

It is also a good idea to have a calculator at hand for converting mile distances to kilometers.

2. What constitutes a georeference

Georeferencing specimen records involves adding coordinates to specimen localities as they are indicated on labels or in catalog books so that they can be mapped in geographic space and used for further analyses and modeling. Specimen localities can be described as qualitative data, while coordinates are quantitative, specific numbers that represent precise locations on the earth's surface. Having coordinates means that specimen records can be integrated with other spatial datasets such as climate, geology and vegetation data. Including georeferences improves the value of specimen datasets significantly as they can be used for a wide range of purposes, such as Red List assessments, distribution modeling, and spatial biodiversity planning.

An increasingly important consideration with biodiversity data like museum and herbarium specimen records is 'fitness for use'. Fitness for use is a measure of data quality based on evaluating the value of a record against a specific proposed use of that record. Records from different sources and datasets are being integrated and published through systems like the Global Biodiversity Information Facility (GBIF) and used for a wide range of purposes, often far beyond the original purposes the data were collected for, by users distant from the originators and holders of the data. Datasets should be comprehensive and of the highest quality possible so that they can be used by a broad audience. Ideally users should not have to refer back to the originating institution or original collectors to determine fitness for use or to resolve data quality questions. Data should speak for themselves.

For this reason georeferences need to be more comprehensive than just a pair of coordinates associated with a locality string. A georeference can be considered an entity in its own right (just like a taxon, a collection event, or a locality) with a set of properties that specify the location, a measure of uncertainty of that location, and additional metadata that assist in interpretation and evaluation of the georeference. The location is represented by a spatial geometry like a point, line or polygon, although points are the most commonly used and simplest to work with. Metadata should include when, by whom, and how the georeference was created. The uncertainty measure is particularly important (although often neglected) as it is the primary means for assessing fitness for use.

The Darwin Core standard for biodiversity data (<http://dwc.tdwg.org>) includes standard properties for georeferences, described in the table below. Any georeferencing projects should be standardized to use these terms. Note that the structure of Darwin Core terms is as whole words and case sensitive - in order for a column heading in a spreadsheet or field in a database to be a Darwin Core term it must match that term exactly without spaces or changes in upper or lower case.

Insert georef terms table here

3. Coordinate formats

Geographic coordinates can be recorded in a number of different formats, for example:

Format	Abbreviation	Example
Decimal degree numbers	DD.DDDDD or DD	-28.63336, 17.01162
Decimal degree string	DD.DDDDD or DD	28.63336°S, 17.01162°E
Degrees and decimal minutes	DD MM.MMM or DM	28° 38.0016'S 17° 0.6972'E
Degrees, minutes and (decimal) seconds	DD MM SS or DMS	28°38'00.1"S 17°00'41.8"E

A variety of symbols can be used for degrees, minutes and seconds. For example, the symbols • ° ' " are all used to represent degrees and ' ' ' for minutes. Sometimes no symbol is used and the terms are written out in part or fully, as in "27deg 15min 45.2sec S 18deg 32min 53.7sec E". Unconventional formats can also be used, for example, degrees, minutes and seconds are sometimes separated by colons, as in "23:26:46.302S 33:56:55.903E". For georeferences the only valid format is decimal degrees as raw numbers, without any symbols or directional (N, S, E, W) indicators. Latitude south and longitude west are indicated with a minus, e.g. '-23.446195, 33.948861' indicates coordinates in the southern and eastern hemisphere. If both coordinates are to be included together in a string then they must be separated by a comma.

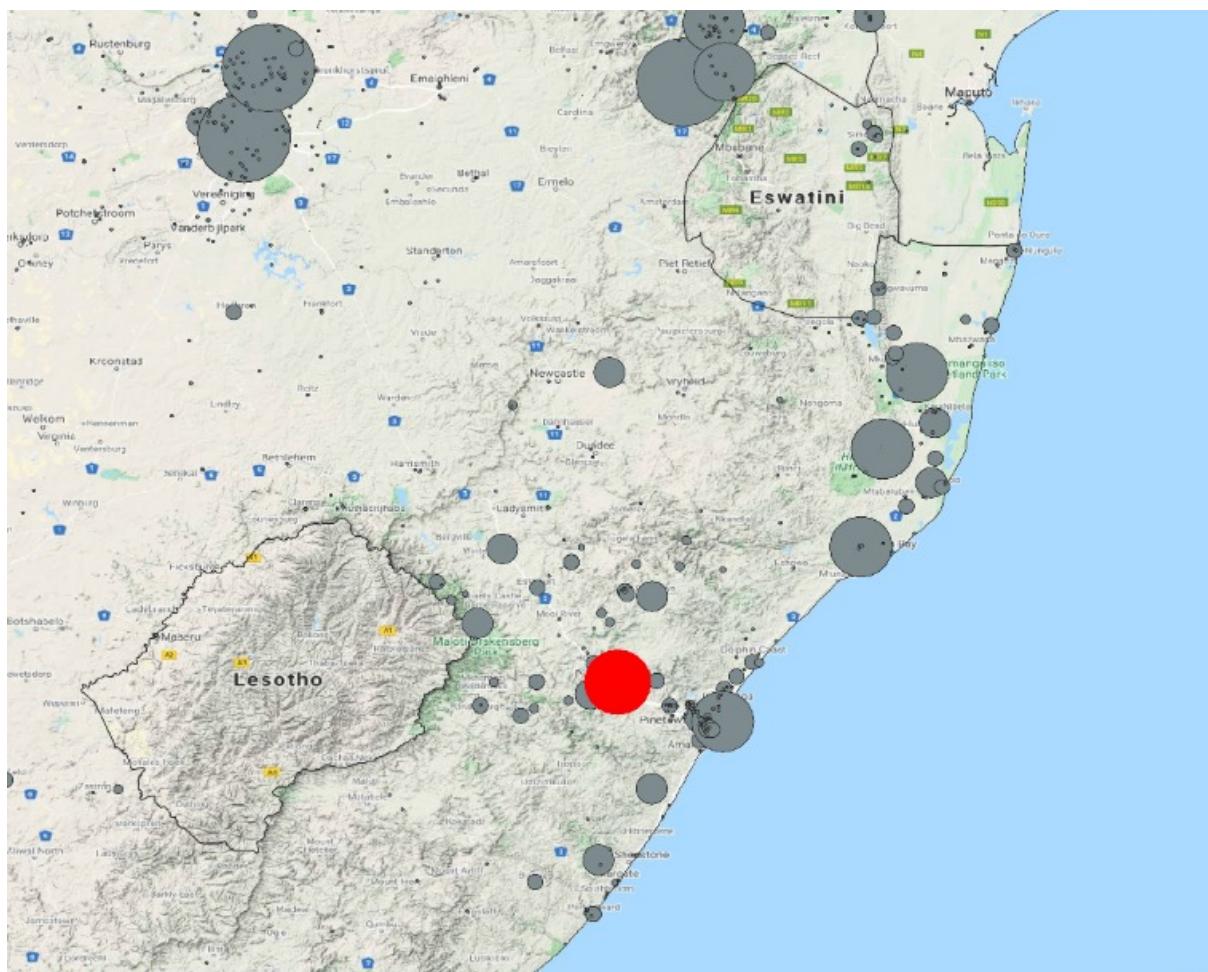
Where coordinates are to be converted from another format to decimal degrees, there are calculators available online. The NSCF also provides an Excel plugin at bit.ly/convertcoords, or the coordinates can be mapped in QGIS using the Lat Long Tools plugin and copied again from the map.

4. Georeference uncertainty

Georeferencing involves two primary steps, identification of the point coordinates of the location described in the specimen locality string, and determining the actual geographic area around that point from where the specimen was most likely collected, termed uncertainty. A georeference is therefore best thought of as a disc placed on the surface of the earth where that disc covers all the possibilities of where the specimen may have been collected. The coordinates indicate the exact placement of the disc, and the uncertainty measure, usually in meters or kilometers, describes the size of the disc.

Geographic coordinates represent a very specific point on the earth's surface. The size of that point is determined by the number of decimal places included in the coordinates. Five decimal places is a point about $1m^2$, approximately the size of a hoola hoop. Six decimal places is the size of a saucer. Seven decimal places is the size of a 20c coin. Because we typically record coordinates in this many decimal places it is important to recognize that those coordinates will not represent the exact point where the specimen was collected. Therefore we need to add a distance measure around that point that represents the total area where the specimen could have been collected. The guiding principle whenever deciding what uncertainty to use is whether the circle represented by the point and the radius capture the specific point on the ground where the specimen was collected.

Assigning this uncertainty measure will depend on the amount of detail included in the locality description. We can describe this as the precision of the locality string. The locality string 'near Upington' is much less precise than '12.2km N of Upington on R360 to Kalahari Gemsbok NP'. The second locality string allows you to more precisely pinpoint the area where the specimen was collected.



This map shows points of varying uncertainty. The greater the uncertainty the larger the size of the circle. The coordinates are placed at the centre of the circle, and the uncertainty measure determines the radius of the circle. These localities are for museum and citizen science records for baboon spiders. The citizen science records have significantly lower uncertainty than museum records because the coordinates are captured at the time of the observation by the observer, and appear as very small circles on the map.

The simplest way to assign uncertainty is to have a set of rules based on the type of locality under consideration, but there are some cases where it can be calculated more precisely. The rules for determining uncertainty are described in the Protocol section below.

5. Protocol

5.1. Locating features

The first step of the protocol is to find the location described in the locality string on a map and assign coordinates to the approximate centre of that location. The NSCF provides a set of GIS data layers for QGIS as well as the SANBI place names Gazetteer to assist with georeferencing. For feature types such as farms or rivers where we have GIS layers the best starting point is to search for the feature in the appropriate GIS layer, using the Select Features function in QGIS. For other locality names you can search in the SANBI Gazetteer in QGIS, Google Maps, OpenStreetmap, and Fuzzy Gazetteer. With practice you will get a sense of which tools work best for finding particular kinds of localities. Remember that there can be spelling mistakes or spelling variances in locality strings, missing or extra letters, and alternative spellings for place names. Examples are

'Gaborone/Gaberone', 'Mafeking/Mahikeng' and 'Buffelsdrif/Buffelsdrift'. The SANBI Gazetteer allows you to conduct a search that considers such potential issues by using spaces in the search strings you enter. For example searching for 'Vin riv' will return 'Vink River', 'Vinkriver', 'Vinkrivier' and 'Vink rivier'. Fuzzy Gazetteer has algorithms for searching for similar place names that consider spelling variations. For example searching for 'Masiene' in Mozambique returns the correctly spelled locality 'Maciene'. Try as many sensible variations of the place name in your search terms when searching using these tools to increase the chance of finding the relevant place names. Also, where there might be different spellings of a place name, search using all possible spellings in Google Maps and SANBI Gazetteer to get all possible options for the locality before selecting one. The best method for finding the location will depend on the type of locality described in the locality string. For example you might use different methods for finding a farm, a mountain pass, or a waterfall. Finding the locality often involves searching through different sources until the feature is found. Features located in Google Maps or the Fuzzy Gazetteer should always be verified against topocadastral maps or the SANBI Gazetteer. To get the centroid of named places like towns use Nominatum (nominatim.openstreetmap.org) so that the centroids are standardized. Coordinates must always be placed at the centroid (best possible approximated centre) of a named place or location.

Importantly, remember that place names do change, particularly in South Africa. Examples include 'Pietersburg' which is now Polokwane, and 'Shabane' in Zimbabwe which has been corrected to the pre-colonial name Zvishavane. If you suspect you have an old name you can do a search for that name in Google to learn more about it and try to work out where the place is situated, and then return to using the geographic sources to get its coordinates. Wikipedia has a good list of place names that have changed in South Africa. The QGIS project provided also includes old maps of South Africa used by John Acocks during his vegetation surveys which include many old place names.

Lastly, this protocol is intended to facilitate rapid georeferencing of large numbers of specimen records from museums and herbaria in a short amount of time. For that reason only the locality information is considered in this protocol. Georeferences apply to the locality, not to the full specimen record. For the bulk of specimen records this is acceptable but it means that in some cases specimens with erroneous locality data will be georeferenced. It is also sometimes the case that the species distribution of the specimen, collectors notebooks, and other such sources need to be consulted to resolve localities. This is again not included in this protocol (see ambiguous localities below) because of the time required. It is recommended that these extra efforts only be taken for important specimens such as types or specimens of potentially threatened species. This protocol also doesn't allow for inference of where a specimen was collected beyond what is provided in the locality string (something which taxon experts are inclined to do). In such cases, where it is necessary, it is better that localities are physically ground-truthed and new, precise locality records taken for whatever the purpose.

5.2. Cultivated or captive specimens

Action: Only georeference if the original wild locality is provided.

Sometimes samples are taken from plants in botanical or urban gardens for preparing herbarium specimens. For animals, captive specimens also sometimes find their way into collections. For plants this is usually indicated with the garden name on the label, or with an indicator like 'cult'. In these cases, only add a georeference if the original, wild locality for the specimen is also included on the label. In these cases, georeference the specimen for the wild locality using the rules below. Otherwise leave coordinates, accuracy, and source field blank, record your name and the date, and indicate in the georeference remarks 'cultivated' or 'captive'.

5.3. The locality value

Action: dwc:locality should have 'oldname [currentname]' if the locality uses an old name.

If you have a locality string and the name is incorrectly spelled or represents an old place name, copy the locality string to a dwc:verbatimLocality field. Then place the correct, new name in square brackets after the locality string in the locality field. For example, if the locality says 'Marrandera', this should be recorded in verbatimLocality and the locality field should be updated to 'Marrandera [Marondera]'. Similarly, if the locality is 'Pietersburg, Tvl', move this to the verbatimLocality and locality becomes 'Pietersburg [Polokwane]'. In this case we remove the 'Tvl', representing 'Transvaal' as Darwin Core uses other fields to record administrative divisions, 'Tvl' is now captured in verbatimLocality, and the Transvaal is no longer a valid administrative division. Spelling mistakes should be corrected in the locality field, but left unchanged in verbatimLocality.

5.4. Districts, regions, cities, large mountain ranges, long rivers, and uncertain localities

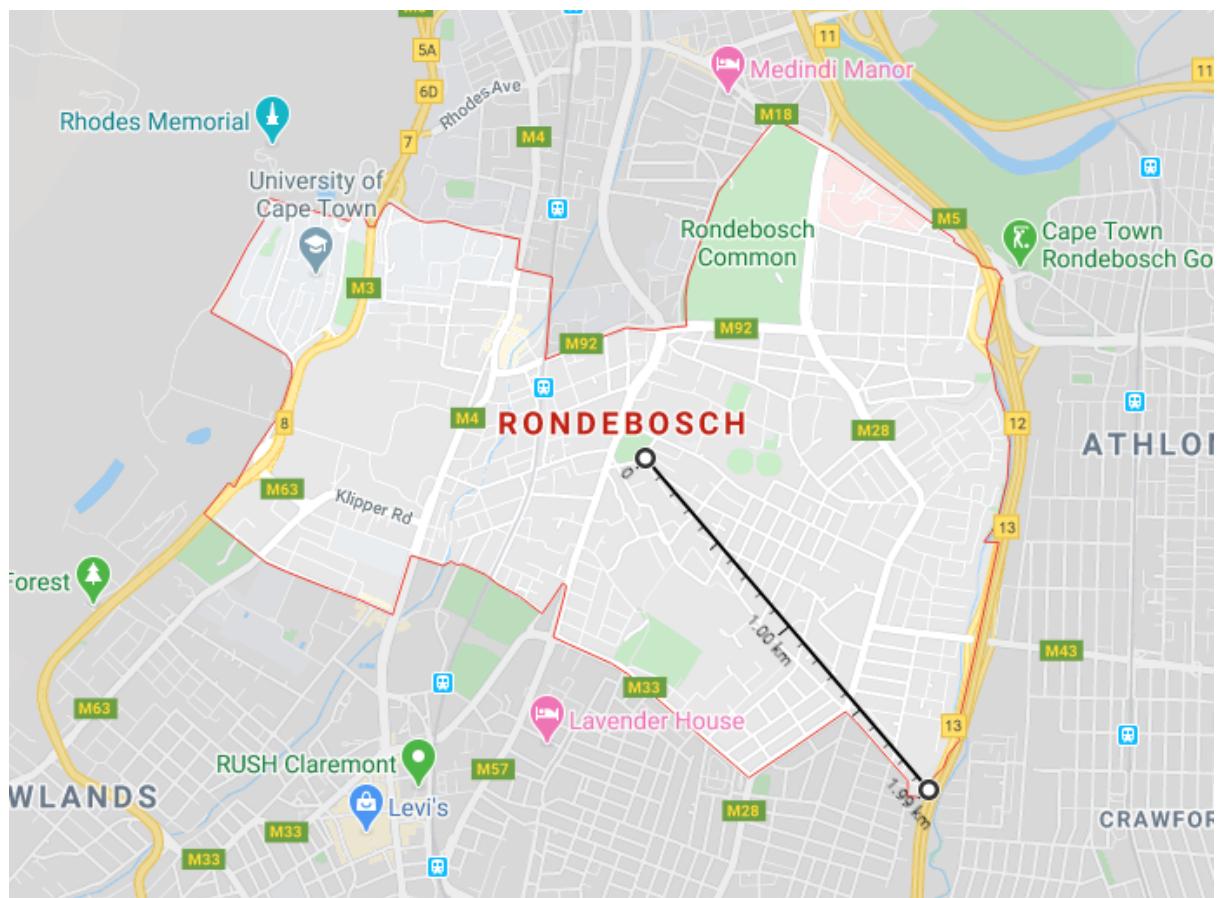
Action: if no QDS then don't add coordinates.

For any locality string that is less precise than a town name or uncertain (usually indicated with a question mark) and no Quarter Degree Square (QDS) is provided, do not add coordinates. These localities are too imprecise to be useful for mapping or analysis. Complete the remaining georeference fields (also leaving uncertainty empty) and add a note to indicate 'too imprecise for georeferencing'. Examples include 'Barberton District', 'Pretoria', 'Magaliesberg', and Bokkeveld, between Calvinia and Vanrynsdorp'. If a QDS is provided, please refer to the section on QDSs below.

5.5. Suburbs

Action: Use Google Maps, point and feature radius.

Google Maps indicates most suburb boundaries. Search for the suburb in Google Maps and then select the coordinates in the approximate centre of the suburb and measure the distance from the centre to the further boundary for the uncertainty measure.



Georeference for suburb Rondebosch in Cape Town in Google Maps.

5.6. Towns

Action: point and standard uncertainty of 10 or 20km.

For localities that only indicate the town name, locate the coordinates at the approximate geographic centre of the town. Nominatum (nominatim.openstreetmap.org) can be used to obtain consistent place name centroids. For larger towns such as Queenstown, Port Nolloth, and Richards Bay (towns with a radius of more than 2km), set the uncertainty to 20km. For smaller towns such as De Rust, Loxton, Wupperthal, or Pilgrims Rest, set uncertainty to 10km. This applies also for locality strings that indicate 'near' the town, without any additional information, such as 'near Queenstown'.

In cases where an approximation is included with a named place, the coordinates are still located at the centroid of a named place. For example, 'Queenstown', 'near Queenstown', and 'Queenstown area', would all have their coordinates at the centroid of Queenstown. Uncertainty is calculated the same way.

5.7. Named places other than towns

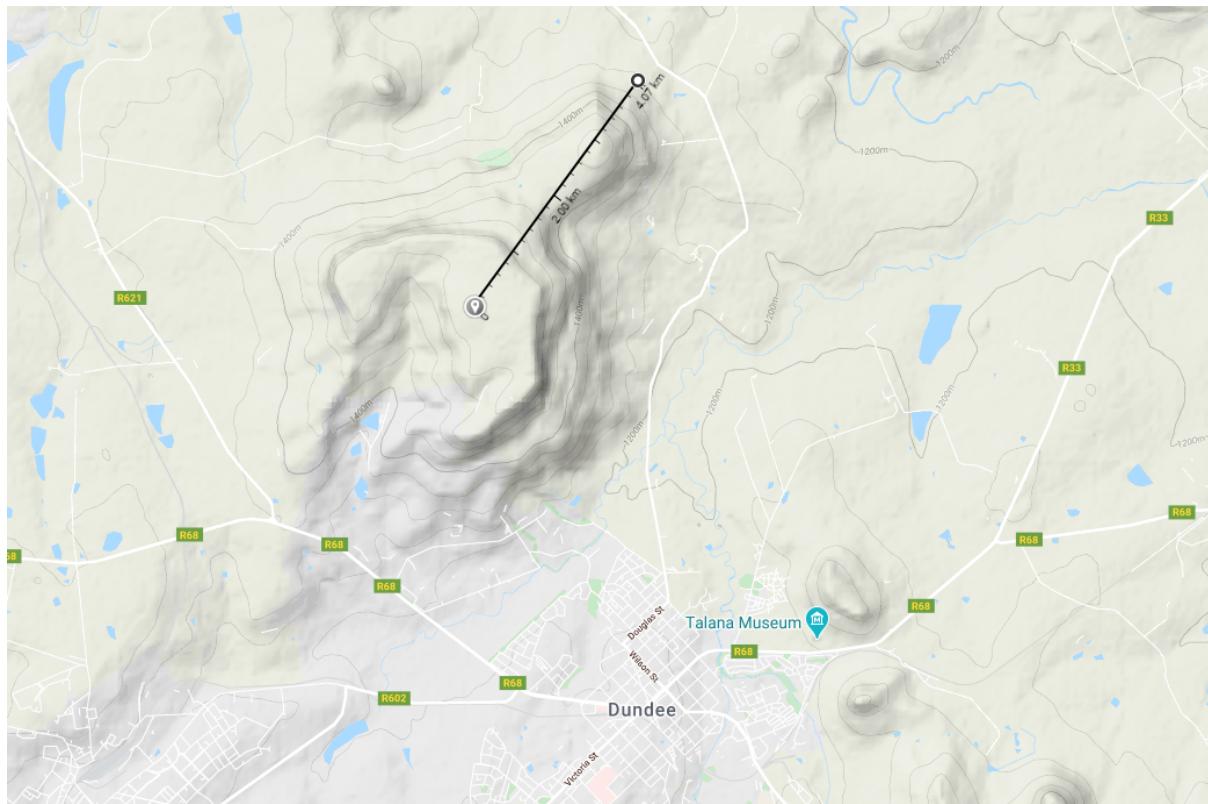
Action: for point locations, use point and 2km uncertainty. For larger locations, use point and feature radius.

For specific named places, such as homesteads, railway sidings, bridges, hills, small villages, dams, river mouths, etc., search for the named place using the various tools you have available such as your GIS layers, gazetteers, and Google. Confirm search results from Google against maps in the GIS. Place the coordinates at the centre of the place. To assign a certainty value, consider whether the specimen would have been collected at that specific place or close to it. If for example the locality

string says ‘Nxwala hill’, you would capture the whole of Nxwala hill with a radius of approximately 1.2km around coordinates placed in the middle of the hill. If it is a very specific location such as a bridge, the specimen is unlikely to have been collected on the bridge, but probably near the bridge. Here an uncertainty of 2km would probably be sufficient to cover the area where the specimen was collected.

Consider that in some cases different places can have the same name (see section on ambiguous locality names below). Also, similar place names might not refer to the same place. For example, ‘Hluhluwe’ and ‘Hluhluwe Game Reserve’ might refer to different places, as there is a town called Hluhluwe as well as a protected area. In these cases you will have to use other information to work out which locality is referred to (see section on ambiguous place names below).

In cases where you have a specific place name and a more general place name, such as ‘Keate’s Drift, near Mpofana’, and you can’t find the more specific place name, find the less specific place name on the map and scan the topocadastral maps near that location for the more specific place name. If you struggle to find the more specific place name and you have a QDS included in the locality string then find the QDS in QGIS using the Select Features function and then scan the topocadastral maps for the specific place name inside that QDS as well as neighbouring QDSs.



Georeference for 'Mpathe hill, near Dundee' in Google Maps, with the radius capturing the entire hill.

Where there are groups or clusters of the same place name in close proximity (a few km) on a map, as sometimes happens around farms or railway sidings, choose the coordinates at the centroid of the cluster of named places and use an uncertainty radius that captures them all.

5.8. Named place with distance and direction from another named place

Action: Georeference the most specific named place using the rules as for named places above.
Distance and direction is just a guideline to find the named place, but increase uncertainty if warranted.

The coordinates must be situated at the most specific place indicated in the locality string and not at the named place itself. Use the same rules as for named places above to place the coordinates and assign uncertainty. Very often the distance and direction from the named place are intended as a rough indication of the location of the locality. For example, for 'Greytown, 20mi NE, KwaMazongwane Forest' the coordinates must be placed at KwaMazongwane Forest, even though in this case KwaMazongwane Forest is 16mi NE of Greytown. Also note that the direction might be imprecise too. For example, the locality string might indicate 'N' of a named place, but the actual locality string could be NE or NW of the named place. This is because collectors often only approximate distances and directions when recording localities. What is important is the most specific named place.

5.9. Distance and direction from a named place only

Actions:

- If 'by road', measure along road, uncertainty depends on the distance value (multiple of 100, 50, 10, 5) measured along the road. Then add named place radius.

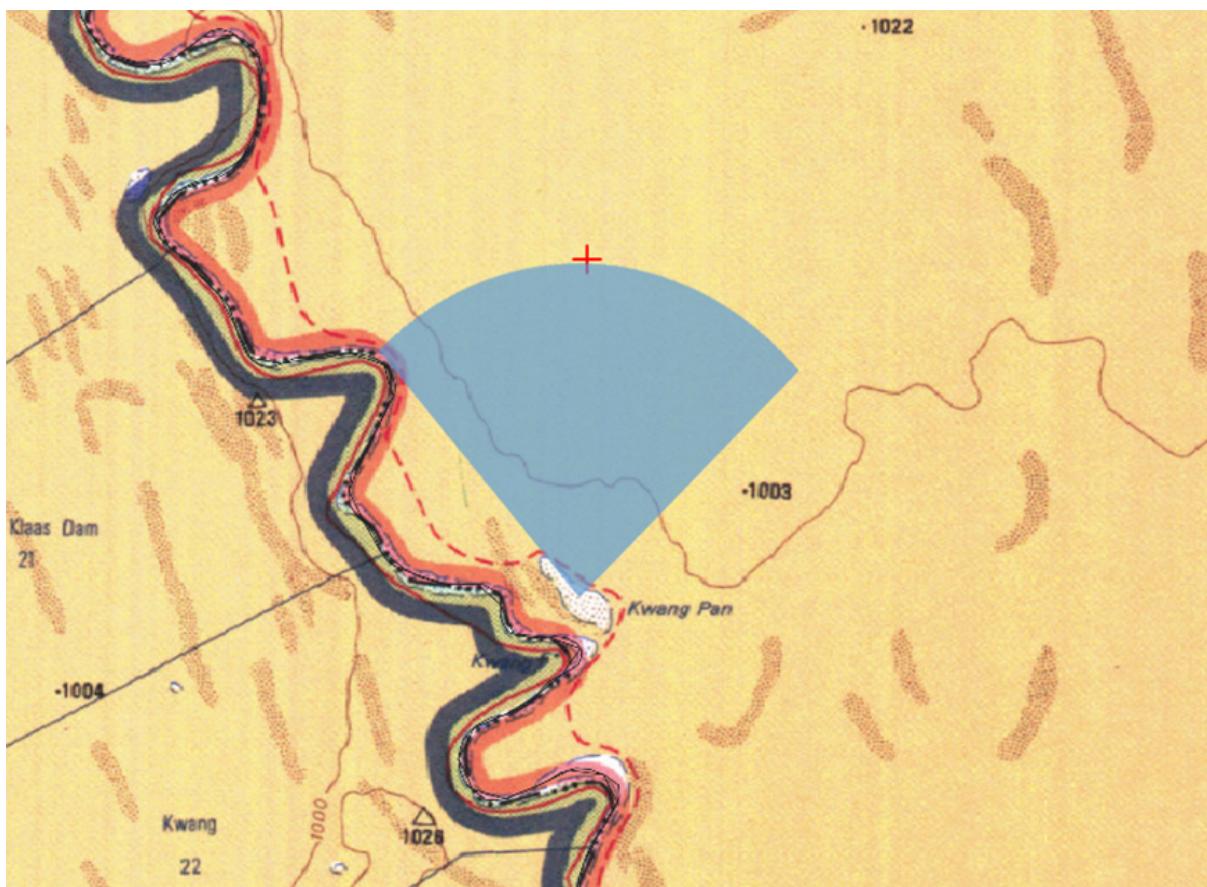
- If 'by air' use georeference calculator for coordinates and uncertainty.
- If neither specified, and there is a road, do both and use point and radius that captures all options.

Many localities are indicated with a distance and direction from a named place, such as 'Kwang Pan, 5mi N'. This indicates that the specimen was collected five miles 'north' of Kwang Pan. If there is no more specific named place included in the locality string then you add coordinates for the distance and direction from the named place. Unless it is stated in the locality string that the distance is by road, you cannot assume that the distance was measured along the road on the basis that there is a road heading in that direction, even though it may be very probable that the specimen was collected along the road. Likewise, unless the locality string says 'due', 'by air', or 'as the crow flies' to indicate that the distance was measured as a straight line, you cannot assume that it was indicated as a straight line. Such localities also seldom indicate where the distance was measured from, i.e. the edge of the place or town or the middle of it. In such cases you need to consider the locations of all these possibilities and place the coordinates in the approximate centre of those possibilities and use an uncertainty value that captures all of them.

Calculating uncertainty for these cases requires several steps because the locality information is quite imprecise. The factors that contribute to the uncertainty are as follows:

- No statement as to whether the distance was measured along a road or as a straight line.
- No indication of where the distance was measured from, i.e. the middle or the edge of the named place, such as a town.
- Uncertainty associated with the direction indicator, as described next.

There is inherent uncertainty in directional indicators as indicated by the various compass directions. N is less precise than NE, and NE is less precise than NNE. N can refer to anything between NE and NW, an angle of 90°. As the distance from the named place increases, so the uncertainty increases because the area between NE and NW gets larger. N, S, E, W are called cardinal directions and all refer to an area of 90° from the named place. The four ordinal directions (NE, NW, SE, SW) have an angle of 45°. The cardinal and ordinal directions together are called the principal winds. We then have eight half winds or secondary-intercardinal directions (NNE, NNW, ENE, WNW, ESE, WSW, SSE, SSW). The half winds have an angle of 22.5°.



An example where compass directions are used, in this case 'Kwang Pan, 5mi N'. The blue area shows the 90° area of uncertainty for the coordinates. In this case there is a track heading NW from the named place, but even though it's most likely that the specimen was collected on that track, we can't assume that. Use the georeferencing calculator for these localities.

The process for georeferencing this type of locality strings is as follows:

If a distance is indicated as 'by road', or the locality string indicates a distance and direction from one named place towards another, like '8 km. SE. of Nelspruit towards Uitkyk', and there is a road from the one named place towards the other then we measure the distance along the road. The starting point for the measurement should be halfway between the middle and the edge of the named place. To calculate uncertainty:

- If the distance unit is a multiple of 100, 50, 10 or 5, then use this value (ie 100, 50, 10, or 5) as the uncertainty, unless the distance equals that value, in which case use half that value for the uncertainty (but see the next step). For example, with '200m N of x by road', the value to use is 100m. With '100m N of x along path' the value is 50m. For '15km N of x by road' the value is 5km. Otherwise, as in '17km N of x by road' the value is 2km. For '5.2km N of x', here the collector is measuring in units of 100m so the uncertainty is 100m.
- Measure this value again from the coordinates you've selected in both directions from those coordinates along the road. We need to do this because distances along roads and straight line distance of the uncertainty measure are often different. Then measure the straight line distance from the central coordinate to the furthest distance along the road. That is the uncertainty.
- If the distance is measured from a large named place, like a town, then the radius of the town needs to be added to the uncertainty.

If the locality string states a direct, straight line distance and direction from a named place, e.g. '10km N by air Giyani', then use the georeferencing calculator to calculate the coordinates and the uncertainty. If the distance measure is a factor of 100, 50, 10 or 5 then round up the uncertainty provided by the calculator to the nearest unit of that factor. Otherwise round up to the nearest single unit. For example, for '10km N Giyani' the calculator provides an uncertainty of 8678m. This should be rounded up to 9km.

If the locality string does not indicate straight line distances or distances by road, then both possibilities by road and straight line directions must be considered. The coordinates should be plotted at the centroid of those possibilities and a radius that captures them all used as the uncertainty.

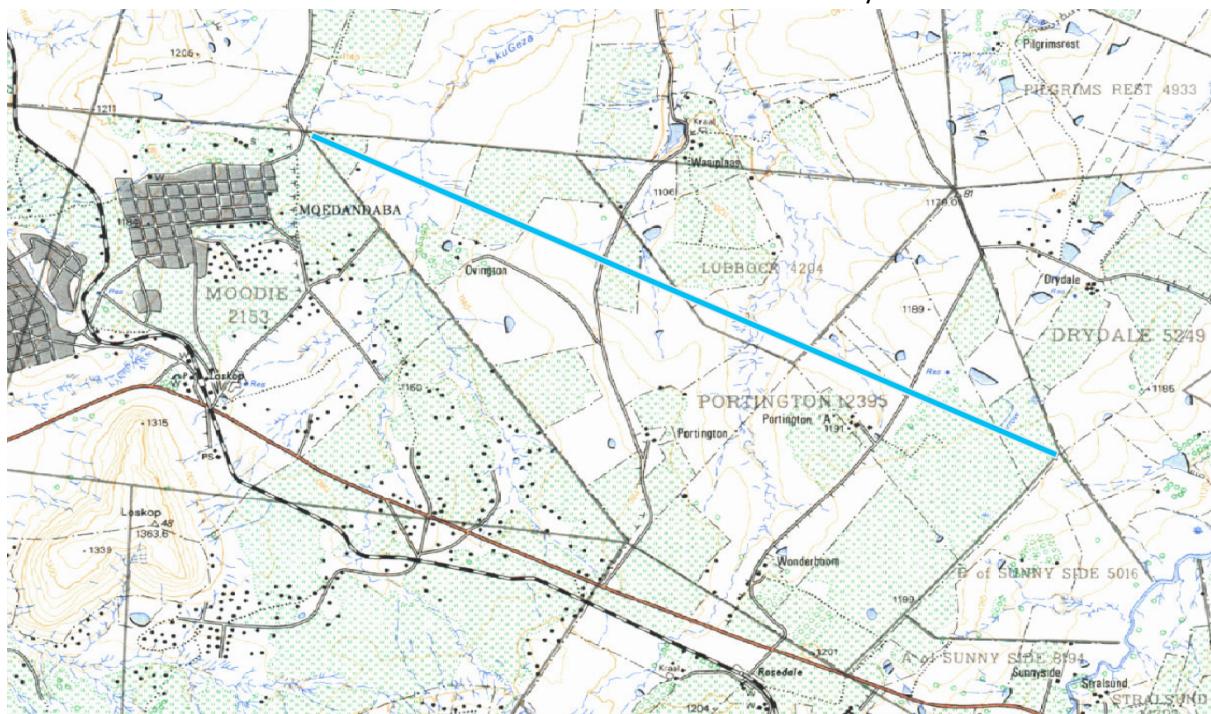
Note that in some cases roads may have changed since the specimen was collected. An example is the N1 highway in Limpopo Province. The original road passed through the main towns of Warmbaths (Bela Bela), Naboomspruit (Mookgopong), and Potgietersrus (Mokopane), and is now the R101. The larger, more direct current route is now the N1. For older records check roads against older maps such as the Acocks maps provided in the NSCF QGIS project.

5.10. Farms and protected areas

Actions:

- **If farm only use centroid and radius**
- **If farm house locate farm building and 1km**
- **If farm gate, find gate and 250m**
- **If near farm gate, find gate and 1km**

Many localities are indicated as farm names or protected areas. Both farms and protected areas vary widely in size, so determining uncertainty on a case by case basis is required. Locate the farm or protected area and measure the distance across its widest portion. The uncertainty is half that distance. In some cases, the boundary of a farm might not be indicated on the available maps. In these cases measure the distance to the nearest indicated farm boundary.



Example with distance measured across the longest portion of a farm. The uncertainty is half of this distance.

With farms in particular, it is important to note that many farms often have the same name. Sometimes the farm number is provided, but it is often not. The farm number helps to resolve the particular farm, however if no farm number is provided then additional information (such as a distance and direction from a town) must be used to resolve the location of the farm. This is an example of why it is important that verbatim locality data from specimen labels or catalog books be transcribed in full and unchanged during data capture processes.

If the locality indicates the farm house or a point near the farm house, locate the primary farm buildings using Google Satellite view or Google Earth and use 1km uncertainty. If there is more than one set of buildings on the farm and it's not clear which is the main homestead, then place coordinates roughly in the middle of the different homesteads and use a radius that captures them all, plus 1km.

When a locality indicates a farm gate, locate the road to the main farm house and place the coordinates where that road enters the farm boundary. It is usually easiest to identify the main farm entry road on Google Satellite view. If the locality says 'near' the farm gate, or similar, use an uncertainty of 1km, otherwise use 250m. If it's not clear where the farm gate is then use the farm centroid and radius and add the additional uncertainty as the specimens may have been collected outside of the farm boundary.

5.11. Mountains, hillsides and mountain passes

Action:

- If 'bottom or top of pass' then:
 - If 'bottom of the pass', and the pass has two bottoms, point between and radius to both bottoms plus 1km.
 - Else point plus 2km uncertainty.
- If just 'pass' then mid-point and radius
- For mountain/hill sides and summits, mid-point and radius.

In addition to the rules for named places above, some additional considerations are important for mountains and hills. Collectors might indicate that the specimens were collected on a particular side of a hill, or the summit of a mountain. In these cases place the coordinates in the centre of that part of the hillside or summit, and measure uncertainty across that whole hillside or summit.

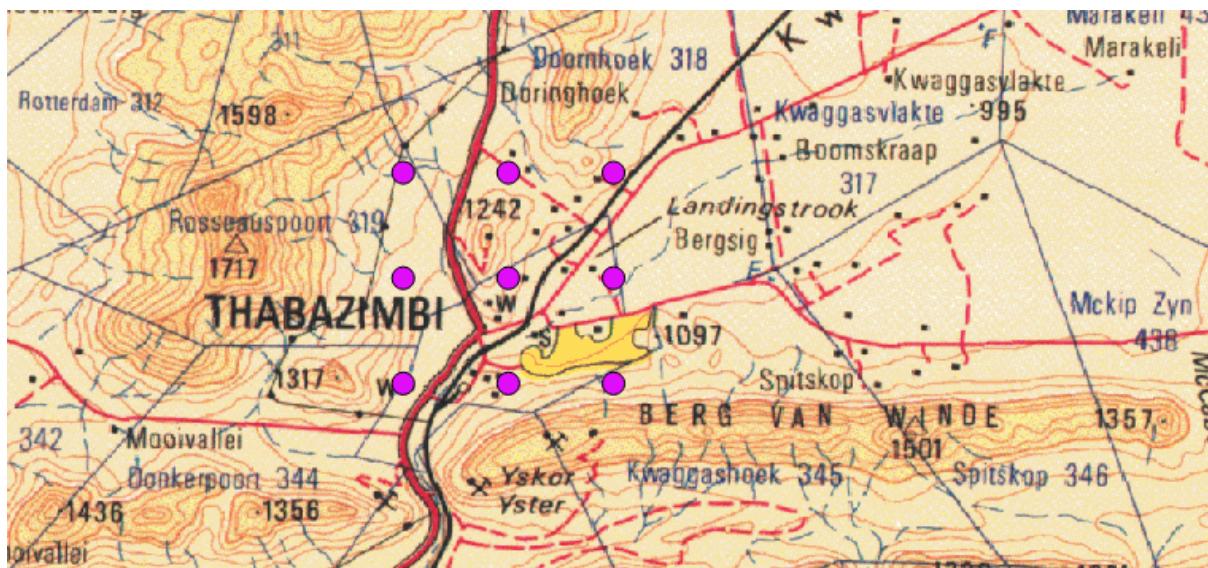
For mountain passes take careful note of what is said about the pass. If only the pass name is provided then place the coordinates at the middle of the pass and measure the straight line distance from there to the end of the pass for uncertainty. Very often though the locality will indicate where along the pass the specimen was collected and is then georeferenced accordingly. A special case is where a locality indicates 'bottom of the pass' and the pass goes over a mountain therefore having two bottoms. In such cases it may be necessary to consult an expert on the species to determine which side it was likely collected on, or you may need to locate the coordinates exactly between the two ends of the pass with an appropriate uncertainty value. Always indicate your choice for these cases in the georeference remarks field. Wikipedia is a good resource for searching for mountain passes and provides coordinates, but those must be mapped back in QGIS and the georeference created using the rules indicated here.

5.12. Localities with coordinates

Action:

- **If degrees and minutes only, georeference in full using the coordinates as a guide**
- **If more precise:**
 - **If there is a QDS and coordinates are QDS centroid with no other locality info, georeference in full using QDS as a guide**
 - **Otherwise plot coordinates in QGIS and confirm they match the locality**
 - **If they match the locality then use coordinates as they are (converted to decimal degrees) with 1km uncertainty.**
 - **If no datum recorded (there almost never is), add 500m uncertainty.**

Sometimes a locality string includes coordinates, which may or may not have been recorded by the collector with a GPS in the field, but these almost always lack a measure of uncertainty or an indication of where those coordinates originated and therefore don't satisfy the requirements of our georeferencing standard. If the coordinates are indicated as degrees and minutes then georeference the locality again using the protocol as outlined above, but use the coordinates as a starting point to find the locality. This is because coordinates with degrees and minutes only are almost always added retrospectively, and are quite imprecise (only around 2km precision).



Presently just for the purpose of the discussion about DM uncertainty, the purple points are a set of coordinates that change by one minute in either direction. Those points are fixed in space. Degrees and minutes can't represent any other point within that area unless decimals are included. The distance between points is approximately 1.6km so the minimum uncertainty possible for DM is +/- 800m. Map scale, lack of a datum and uncertainty as to whether the coordinates are provided by the collector or added retrospectively all increase the georeference uncertainty further.

If the coordinates are as decimal degrees, degrees and decimal minutes, or degrees minutes and seconds, then plot the coordinates in QGIS and check that the location they indicate matches the locality string. If not then it is most likely the result of a coordinates transcription error during data capture, but GPS malfunctions and other errors are also possible. Georeference the locality again using the rules above, then compare your georeferenced coordinates to the specimen coordinates. This might indicate whether there was a transcription error. If a transcription error is detected then correct these and use the corrected coordinates and determine uncertainty as below. If not, then use the new georeferenced coordinates and indicate in the remarks that the verbatim coordinates appear incorrect.

Once the verbatim coordinates have been corroborated or corrected, use these as the coordinates for the georeference after converting to decimal degrees and set the source field to 'verbatim coordinates' and add an uncertainty of 1km. This may seem excessive but without a specific indication of uncertainty from the collector there is a risk of implying false precision with a lower uncertainty value. Real accuracy can range from a few meters in cases where the collector records the coordinates at the specimen in the field, to 100s of meters when collectors take a single coordinate for the whole area they visit in a collecting event. The same collector can also do different things at different sites depending on the purpose of the collecting. Where needed, as in the case of red list assessments, individual records can be scrutinized and uncertainty values updated in consultation with collectors or other knowledgeable collections staff.

In addition to the uncertainty above, unless there is an indication of the datum that the coordinates were recorded in, add another 500m to the uncertainty. This is because coordinates recorded in different spatial reference systems refer to different places on the surface of the earth and in southern Africa the commonly used datums (Cape, Clarke 1880, WGS84 and Hartebeeshoek) can be up to 500m different from one another.

If a QDS is provided, and the record includes coordinates, make sure to check if those coordinates are not at the centroid of the QDS. If they are then ignore those coordinates and georeference the locality again following the rules of this protocol.

See the section on Quarter Degree Squares (QDSs) below for how to handle cases where a locality includes coordinates as well as a QDS.

5.13. Quarter degree squares (QDSs)

Action:

- **Use QDSs to start finding place names. Search surrounding QDSs as well**
- **If QDS and area like protected area, and QDS falls completely or almost completely within the named area, use the QDS centroid**
- **If QDS intersects with larger area like a city or mountain range AND date collected is after 1970, point and radius for the section within the QDS**

Sometimes there will be a QDS included with the locality string. QDSs may have been recorded by the collector while in the field, taken from topocadastral maps or calculated from coordinates provided. This uncertainty as to their origin means that QDSs need to be treated with some caution in the georeferencing process. The QDS system was put in place in the early 1970s, so any records earlier than that including a QDS must have had it added later.

If there are no coordinates provided, then use the QDS as a starting point to find the most specific place name indicated in the locality string. If you cannot find it within the named QDS, then search the surrounding QDSs. Uncertainty is calculated as normal using the rules for the type of locality. If there is no other locality information other than the QDS then use the QDS centroid of the coordinates, with 20km uncertainty.

If you have a place name that refers to a large area, such as a city, farm or a mountain, as well as a QDS, then use the area of the feature within the QDS as the area to georeference. For example, if the QDS intersects with a mountain, place the coordinates at the centroid of the part of the mountain that falls within the QDS and use the radius of the part of the mountain within the QDS as the uncertainty. Do the same for other feature types.

5.14. Using habitat information and elevation

Action: use habitat to refine location if possible, but be careful.

Sometimes locality strings will include topographic or habitat information that might be helpful in assigning accurate coordinates. For example, if the locality string includes 'bottom of mountain', the bottom of the mountain can be located on topocadastral or terrain maps quite easily. On the other hand, habitat information might also not be helpful in assigning coordinates. For example, 'bush clumps next to the side of the road' could refer to any bush clumps along the road, including bush clumps that existed at the time of the collection and don't exist now. When deciding whether to use habitat information in allocating coordinates the principle should be to use permanent fixed features only such as mountains, rivers, and dams, and to avoid using impermanent features like patches of vegetation. An exception might be forest patches which are relatively permanent. Avoid using geology or soil types in making georeferences as outcrops of particular rock types are not accurately mapped on most geological maps.

Elevation is sometimes provided with locality or habitat information for specimens collected on mountains and may be helpful in assigning coordinates accurately, but should be used with care if not accompanied by an indication of aspect or specific side of the mountain where the specimen is collected (see calculating uncertainty below).

5.15. Ambiguous or contradictory localities

Action: don't add coordinates and state 'ambiguous' in remarks

For ambiguous place names, i.e. those that can refer to multiple places, you need to use additional information to try and resolve the correct location where the specimen was collected. Examples include 'Nooitgedacht' and 'Buffelsdrif(t)'. If there is no additional locality information in the locality string, and you only have the ambiguous place name, then you might need to look for other localities sampled by the collector at around the same time as this locality was sampled. This may require additional work and it might be better to set these examples aside for georeferencing separately in favour of keeping momentum with the more precise localities. You might also need to consult a taxonomic expert to narrow down the possibilities based on the distribution of species collected at the locality. Beware of examples like 'Mpumalanga', which might seem obvious but could refer to the province or to the township near Durban. There are also examples where different towns in different provinces have the same name, like 'Heidelberg' (Gauteng and Western Cape) and 'Middelburg' (Mpumalanga and Eastern Cape). Also be careful to consider cases like 'Hluhluwe' which might refer to the town of Hluhluwe or Hluhluwe Game Reserve. In these cases don't assume it's one or the other. If you select one over the other make sure to include the justification in the georeference remarks field.

Where a locality string includes two possibly contradicting place names, first consider whether one might refer to a broader region, and the second to a more precise location within that region. For example 'Komatipoort, Barberton', can be taken to refer to the specimen/s being collected at Barberton, which falls within the Komatipoort magisterial district. If however, the relationship between the place names cannot be resolved and they appear contradictory leave the coordinates blank, complete the remaining georeference fields and indicate in the notes that the locality cannot be resolved.

6. Georeference sources

It is important to record the maps and other resources used to identify the location of the coordinates used for the georeference. For example, if you look up a place name in the SANBI gazetteer and then verify it on the 1:250 000 topocadastral maps then both the gazetteer and the topocadastral maps must be recorded in the source field. When recording more than one source use a pipe character “|” to separate them. Abbreviated names for the sources, such as ‘1:250k topo maps’ are acceptable as long as it is clear what they refer to.

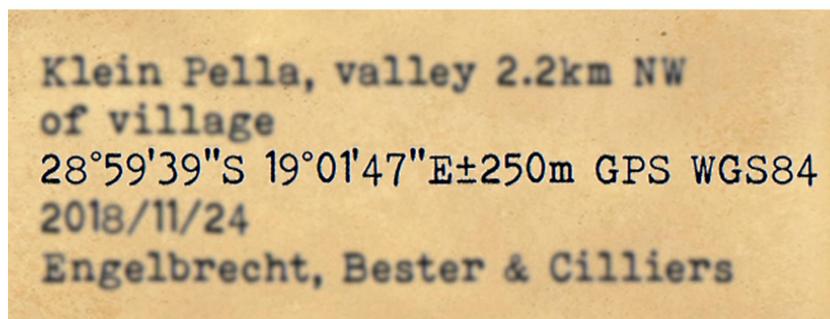
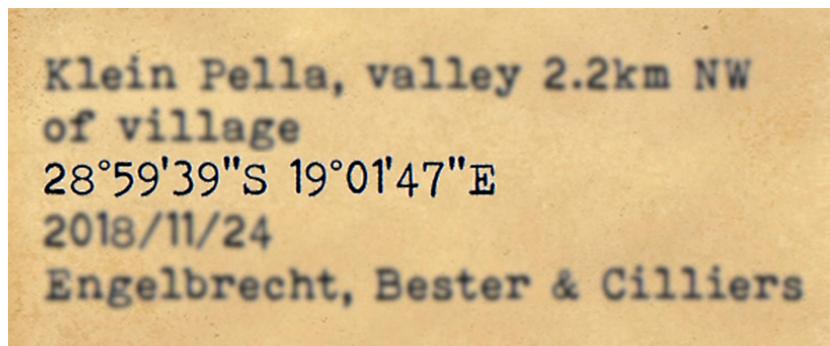
7. Georeference remarks

The remarks field is very important during georeferencing and should be used to record anything important about how a georeference was made, particularly any assumptions made in the placement of the coordinates. Any departures from the protocol used must also be noted.

8. Completing the remaining fields

After adding the coordinates and uncertainty complete the remaining georeference fields. These fields provide record-level metadata on when and how the georeference was made and are

important for evaluating fitness for use of the data. See the table on the Darwin Core georeference terms above for what should be included in each of the fields. Make sure to indicate the version of the protocol followed if it is a versioned protocol such as this one. In general, georeferencing should be done using the WGS84 datum and this should be recorded in dwc:geodeticDatum. The datum for the QGIS project provided by the NSCF and Google Maps is WGS84. WGS84 should also be used as the standard datum for field data collection using GPS, and should be indicated on specimen labels and in collection databases.



A comparison of specimen labels showing coordinates with no supporting information, leading to increased uncertainty, and coordinates with supporting information leaving no doubt as to their certainty. Ideally collectors should always provide the supporting data indicated on the second label.

9. Common abbreviations in location strings

- btw. = between
- C. Col. = Cape Colony (approximately all the current Cape provinces)
- Distr. = District
- Div. = Division
- F.R. = Forestry Reserve
- Ft = Feet
- Ftm = Fathoms
- Gt = Great
- int. = intersection
- k, km = kilometers
- m = meters OR miles
- mi = miles
- Mt, Mtn, Mts = Mountain/mountains
- nr. = near
- N, S, E, W, NE, SW, etc = these are compass directions, north, south, etc.
- Nat. Res. = Nature Reserve
- N.P. = National Park
- prob. = probably

- Pt. = Point
- Riv, riv. = river
- St = Street or Saint